SAMPLE SELECTION

SAMPLING SPECIFICATIONS

SAMPLE SELECTION

This section explains in detail the procedures for selecting the samples for TPS review. Information is provided regarding the requirements of a sampling system, options available for selecting samples, and details for each function being reviewed by TPS. Flexibility of **options** has been included in the procedures explained. Each State should select the option best suited to their particular operation for each tax function. The option preferred by the N.O. will be indicated with reasons for the recommendation.

WHAT DOES SAMPLING REQUIRE?

The TPS sampling methodology for all the tax functions contains five distinct steps.

1) **Identify, find, or gather data elements for sampling Universe files;**
2) **Extract or collect data to create the Universe files;**
3) **Determine which transactions to select for the sample;**
4) **Select the cases to review;**
5) **Create output reports and files of the selected cases.**

This process is illustrated in Exhibit A - 1.

How these five steps are accomplished is the SESA's choice. The design of TPS allows considerable flexibility in the method used to select sample cases. Not all States have the same level of automation,

and, even within a State, not all aspects of the tax program are computerized. Varying file structures may lend themselves to different sampling approaches.

SAMPLING SPECIFICATIONS

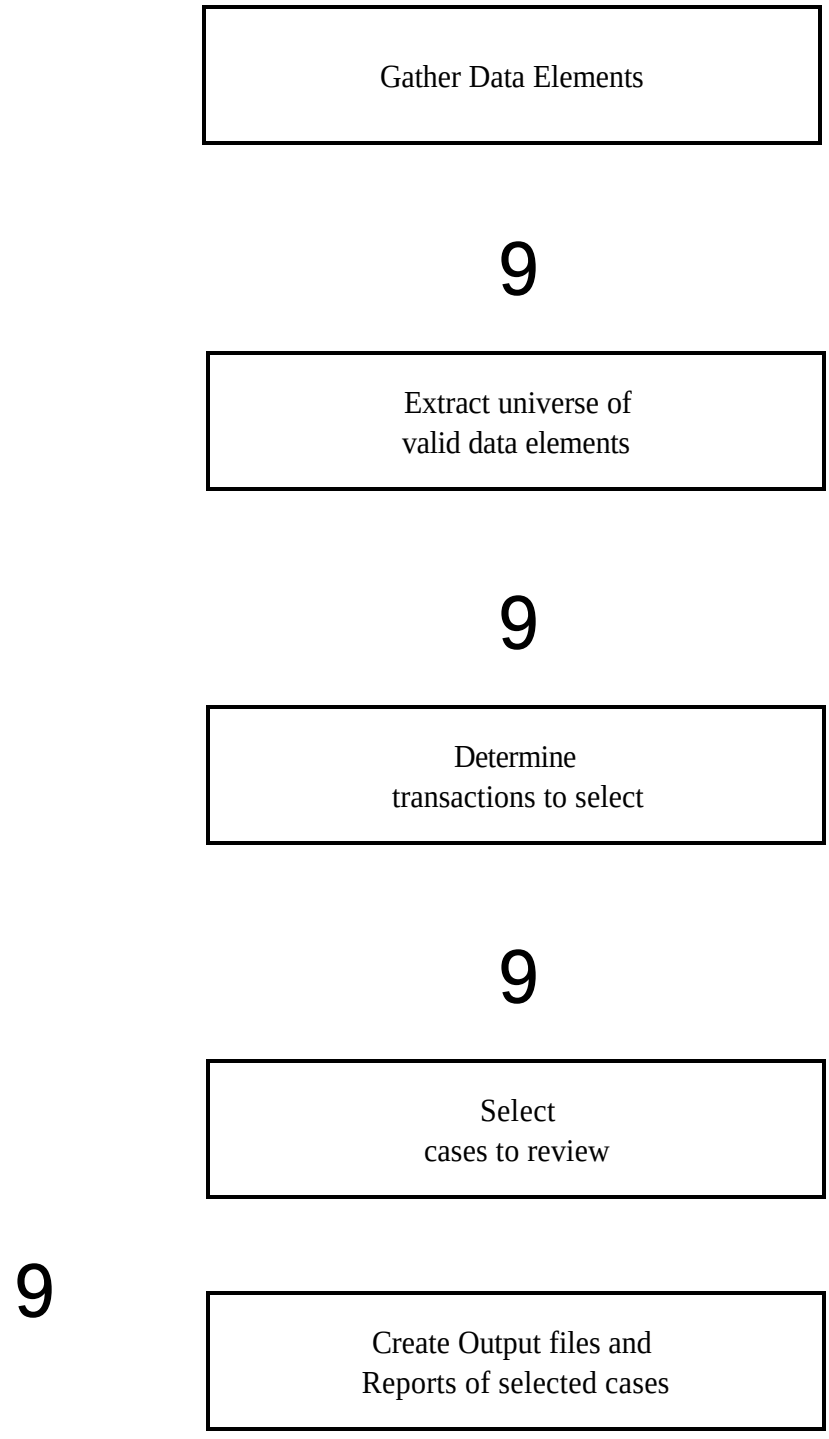
SAMPLE SELECTION

If files or sources of information are automated, **the TPS reviewer must work with the ADP staff to determine the best sample selection method for <u>each</u> of the <u>fourteen</u> samples required.** The same sampling method does not have to be used for all functions. This decision will hinge on the level of automation, or types and location of files involved.

If files are kept manually, the TPS Reviewer must research the sources of information needed to create the Universe.

The SESA may use an automated or manual method to create the universe files and select the samples. Throughout this section of the Appendix, both methods will be discussed. Various details and tips on how to decide which is the best approach for your State will be provided.

5 TPS SAMPLING STEPS



SAMPLING SPECIFICATIONS

SAMPLE SELECTION

Step 1. - Gather Data for the Sampling Universe

Collect Required Data

The first step is to gather or have access to **THE UNIVERSE (ALL)** of the particular transactions to be reviewed. Transactions can be inputs such as status determinations, outputs such as experience rate notices or a condition at a point in time such as an account with monies due. **It is essential that every transaction or item meeting the criteria be included.** This means that all possible sources or locations of the transactions must be searched. For example, information for completed field audit cases may be kept at the central office as well as in field locations.

Be sure to check that only valid transactions are included.

For instance, only Tax Rated employers are included in the universe for the Employer Experience Rating sample.

Refer to the definitions in the TPS Handbook under the "UNIVERSE" section of the Acceptance Sample Instructions for each tax function.

Depending upon the automation level of the SESA, the creation of the universe may require the extraction of data from a computer (mainframe or PC) or the keeping of a manual list or searching file drawers.

SAMPLING SPECIFICATIONS

SAMPLE SELECTION

If the automated sampling method is chosen, the SESA ADP staff is responsible for creating universe files which

contain the requested information. **The TPS Reviewer, however, must work with the SESA ADP staff to determine which of the data elements on the SESA's computer system are needed to create the employer transaction records that belong in the Universe.**

The National Office has developed specifications of the minimum data needed for each sampling universe. (Refer to the file layouts in the sampling instructions for each tax function). The TPS reviewer may request additional data elements to include in the sampling universe to facilitate the review process, perform special studies, or to use for exception samples (refer to section IV, Appendix A). The SESA ADP staff is responsible for creating the programs and/or utilities to extract or gather the requested data elements.

Use Proper Time Period

Each sample being reviewed will be selected from a universe covering a **specific time period**. For example the field audit sample is from a universe made up of all audits completed during an entire calendar year.

Other samples cover activity in a particular quarter or at a specific point in time. Read the instructions carefully for each tax function's samples to be sure the correct time period is used.

SAMPLING SPECIFICATIONS

SAMPLE SELECTION

Once the elements for the universe files have been identified, and the extraction program created, it is recommended that the TPS reviewer examine a small cross section of the records (approximately 50 -100) to verify that the data elements are correct and the proper time frames are being followed. **It is strongly recommended that the data verification of the universe files be accomplished well before the actual time period that the data is needed and the actual universe files are created.**

After the TPS reviewer has approved the data elements and time frame, the SESA ADP staff should establish procedures for building the universe files, selecting the sample cases and saving the universe files.

Schedule Data Capture

We strongly recommend that in building the universe files most transactions be captured as they occur.

This is important since the desired transaction may be superseded by an subsequent transaction and the desired information is no longer readily available.

Status Determinations amply illustrate this scenario. For example, an employer is determined to be a NEW employer, then it is discovered that he is in fact a SUCCESSOR employer. Before the end of the year, the employer goes out of business and is TERMINATED. All three determinations occurred in the same

year and should be included in their respective universes. However, in some systems, the information for the NEW determination could be replaced by the information for the SUCCESSOR determination and then the SUCCESSOR determination is replaced by the information for the TERMINATION determination. Any trace of the NEW and SUCCESSOR determinations, is effectively lost.

This can be especially true in highly automated States where data fields are often overlaid with the most recent information.

Some States may be able to reconstruct events by using daily transaction logs maintained in their data processing environment. For the most part it is still better to capture TPS transactions as they occur during the time period to be reviewed.

In States with manual systems, the transaction source information may be filed in a filing cabinet by employer identification number, meaning the transactions for the time period needed by TPS are mingled with those occurring at other periods of time. Some suggestions to alleviate this problem:

- apply special labels to the files;
- store files in separate file drawers;
- keep a manual list before filing; or
- maintain a separate fiche

SAMPLING SPECIFICATIONS

SAMPLE SELECTION

The Universe should be isolated in some manner so that the integrity of the data will not be compromised when it is time to select the sample.

SAMPLE SELECTION

Step 2. - Collect Data to Create the Universe

Data Collection Automated Approach

Each SESA has the option to collect the data for the universe files any way they choose. If the particular tax function is automated, the SESA ADP staff should create a program to gather the data. This program can be written using any computer language or utility such as Easytrieve, FileAid, or SAS. The N.O. recommends that the SESA follow the N.O. file layouts provided for each tax function. The resulting transaction files can then be used as input into the N.O. Sample Determination program.

Data Collection Manual Approach

For tax functions that are not automated, the SESA must collect the information manually. This can be accomplished many ways but a simple method can be the creation of a paper list.

For example: as Field Audits are completed, each auditor will write the employer identification number and the date the audit was completed on a paper list. Throughout the year, the list is updated possibly onto numerous sheets of paper. At the end of the year, all the lists are assembled as the Universe and the number of completed audits are sequentially numbered and counted. All possible locations of the data MUST be included in the data collection process.

SAMPLING SPECIFICATIONS

SAMPLE SELECTION

Save Universe Files

It is important to save the original universe files. States **MUST** save the universe files from which samples are selected even after providing the selected cases to the TPS reviewer. The reasons for this are:

- a second acceptance sample may be needed;
- an expanded sample may be needed;
- a particular case may have to be replaced; or
- a special study is needed.

In each instance above, the original universe file would need to be used again.

SAMPLING SPECIFICATIONS

SAMPLE SELECTION

Step 3. - Determine Records for the Sample

Perform Calculations The third step is to determine which records to select for the sample. The formulas used to determine which records to select must be the formulas provided by the N.O. The calculations performed are designed to ensure a non-biased systematic sample. See Exhibit A - 2 for these formulas.

To perform the calculations, three numbers are needed:

- **Total Records in the Universe** - Once the universe has been created, a count of all the transactions in the universe must be performed. This count is represented by "P" in the calculations.

- **Number of Records to Sample** - The number of cases to sample is dependant upon the size of the universe. In most cases, the size of the acceptance sample is 60 cases. However, occasionally the universe is so small that a smaller sample size is used. If the universe is below 1,200 cases, a smaller sample size is necessary. See the chart on the following page to determine the sample size needed for populations below 1,200. Sample size is represented by "N" in Exhibit A - 2.

- **Random Number** - This is the third critical number necessary to perform the sample calculations. It is represented as "R" in the formulas. These numbers will be supplied by the N.O. An example is provided in Exhibit A - 3.